

rates within an RDD sample produce a disproportionately high number of college graduates and a disproportionately small number of high school dropouts. To correct for differential rates of participation in the interviews, an initial **90-cell** weighting matrix was used that includes five age strata, three **racial-ethnic** strata, two gender strata, and three educational strata. The cells used in the development of the weight match the cells presented in the under-coverage table (see Table 8). Estimates for the U.S. population were obtained from the Bureau of the Census' *Current Population Report*<sup>6</sup>. The 90 cells were then collapsed into 52 cells to control for cells in which only a small number of respondents occurred. The weight variable **WT5** should be used in all analyses producing population estimates for the United States. (See Appendix G for the final weight matrix used for the study.)

## **DATA EDITING AND FILE CONSTRUCTION**

Professor Miller and Dr. **Kimmel** were responsible for the primary data editing and file construction tasks. The following sections describe the major file construction tasks conducted for the 1999 *Science and Engineering Indicators* study.

### **Data Editing and Processing**

The survey data that were collected were monitored throughout the data collection period. After the first few days of data collection, a preliminary set of data was prepared and placed on the NORC web site, along with the SAS commands that could be used to read the data. Dr. **Kimmel** downloaded the data and created an SPSS analysis file. Cross-tabulations were run to ensure that all skip patterns were operating correctly. Dr. **Kimmel** also produced a listing of all open-ended responses, and examined them for any potential problems. A few minor problems were identified with interviewer entries of a few open-ended questions, and these problems were reported to the NORC staff.

Dr. **Kimmel** retrieved the survey data from the NORC web site on a weekly basis after the initial data reviews. Each week Dr. **Kimmel** constructed an SPSS analysis file, prepared a preliminary listing of frequencies for all of the variables, and examined the frequencies for any potential problems. The reviews of the frequencies focused on such things as values outside the normal or probable bounds (eg., a respondent's age being listed as under 18 or over 110); values not specified in the questionnaire (eg., a value of 5 appearing for a true/false question, in which only the values of 1, 2, 8, and 7 were expected); as well as unusual patterns of responding based on prior survey results (eg., all respondents indicating that they had a clear understanding of a molecule, or no respondents indicating that they read a newspaper every day). No problems were identified in these reviews.

### **Coding of Open-ended Questions**

The 1999 study included two types of open-ended questions: short answer open-ended questions and long answer open-ended questions. Short answer open-ended questions in the 1999 study include such items as occupation, magazines, college major, and subjects searched for on the World Wide Web. Long

---

<sup>6</sup> See Series P-20, No. 462, *Educational Attainment in the United States: March 1991 and 1990*, U.S. Government Printing Office, Washington, D.C., 1992.

answer open-ended questions included in the 1999 study were the meaning of various scientific terms such as DNA, molecule, and radiation. Short answer open-ended questions were coded on an on-going basis, while long answer open-ended questions were coded in three batches, after 1,000 completions, 1,500 completions, and 1,882 completions.

Each week Dr. **Kimmel** prepared files containing the responses for all short answer open-ended questions collected during the previous week. The Electronic Coder (EC) software program was used by Research Assistants to code all short answer open-ended questions. Short answer questions were coded in a relatively mechanized **manner** by identifying key words in the respondents' answers. After the first week of coding, the Research Assistants provided Dr. **Kimmel** with hard copies of the codes and actual responses for each open-ended question. Dr. **Kimmel** reviewed the codes with the Research Assistants and made changes where needed in the codes assigned, and provided additional training as needed in the use of the EC and the codes to be assigned. The codes assigned to the short-answer questions were merged with the main SPSS analysis file on a weekly basis. At the conclusion of the study Dr. **Kimmel** reviewed all of the codes assigned for each short answer open-ended question.

Questions classified as long answer open-ended questions varied in length between 74 and 560 characters. These responses cannot be coded using the EC, for they require complex judgments by the coders. Three graduate students in biology and chemistry were hired to code the long answer open-ended questions included in the 1999 study. All of the graduate students had also coded some of the same open-ended responses for two previous surveys conducted by the International Center for the Advancement of Scientific Literacy. Dr. **Kimmel** reviewed the coding categories with the coders at the beginning of the survey, and provided the coders with written coding instructions, and examples containing responses assigned each possible code from past surveys for each of the questions. The coders worked independently, and were not allowed to discuss their coding assignments with the other coders. At the conclusion of the survey, at least Dr. **Kimmel** reviewed all inter-coder disputes (cases in which the three coders were not in agreement) and assigned a final code to those cases.

### **Construction of Scales and Other Summary Measures**

A number of constructed variables were created at the conclusion of data collection for use in the preparation of the appendix tables for ***Science and Engineering Indicators***. The major constructed variables include attentiveness to selected public policy issues, exposure to formal science and mathematics education, and various categorizations of the respondents' occupation, education, and age. (See Appendix I for the SPSS commands used to create all constructed variables.)

## **QUALITY PROFILE**

### **Review of Questions**

Several pieces of information were included with each case to assist with the review of data quality. First, interviewers were asked to rate each respondent's comprehension of the questions included in the study, by assigning a code of high, moderate, or low to each case. A total of 113 cases, or 6 percent of the sample, were assigned a low comprehension rating by the interviewers. Dr. **Kimmel** reviewed each